Joint prediction of multiple scores captures better individual traits from brain images

Mehdi Rahim, Bertrand Thirion, Danilo Bzdok, Irène Buvat, Gaël Varoquaux

Joint prediction of multiple scores captures better individual traits from brain images

Mehdi Rahim[a], Bertrand Thirion[a], Danilo Bzdok[b], Irène Buvat[c], Gaël Varoquaux[a]

[a]*Parietal Team - Inria / CEA - Paris Saclay University, France*
[b]*Department of Psychiatry, Psychotherapy and Psychosomatics, RWTH Aachen University, Germany*
[c]*IMIV group - Inserm / CEA / Univ. Paris Sud - Paris Saclay University, France*

Abstract

To probe individual variations in brain organization, population imaging relates features of brain images to rich descriptions of the subjects such as genetic information or behavioral and clinical assessments. Capturing common trends across these measurements is important: they jointly characterize the disease status of patient groups. In particular, mapping imaging features to behavioral scores with predictive models opens the way toward more precise diagnosis. Here we propose to jointly predict all the dimensions (behavioral scores) that make up the individual profiles, using so-called multi-output models. This approach often boosts prediction accuracy by capturing latent shared information across scores. We demonstrate the efficiency of multi-output models on two independent resting-state fMRI datasets targeting different brain disorders (Alzheimer's Disease and schizophrenia). Furthermore, the model with joint prediction generalizes much better to a new cohort: a model learned on one study is more accurately transferred to an independent one. Finally, we show how multi-output models can easily be extended to multi-modal settings, combining heterogeneous data sources for a better overall accuracy.

1. Introduction

Between-subject variability, be it in image-based brain features, psychological tests or clinical assessments, is a window into human neurosciences. Relating behavioral assessments to brain images helps grounding inter-individual variability in anatomical and functional aspects of brain organization (Dubois and Adolphs, 2016; Abi-Dargham and Horga, 2016). The corresponding imaging-based biomarkers can then serve as intermediate phenotypes, or *neurophenotypes* to characterize the subjects (Drysdale et al., 2016). In this context, recent years have witnessed the emergence of large-scale neuroimaging studies with rich assessments of subjects' traits and multiple brain-imaging modalities. Such studies aim at characterizing healthy subjects, such as the Human Connectome Project (Van Essen et al., 2013), investigating brain genetics, *e.g.* Enigma (Thompson et al., 2014), or the impact of ageing, *e.g.* LIFE (Loeffler et al., 2015) or AIBL (Ellis et al., 2009). They also investigate neurological pathologies as Alzheimer's Disease in ADNI (Jack et al., 2008), psychiatric disorders, as autism in ABIDE (Di Martino et al., 2014), or prospective epidemiology, as with UK biobank (Elliott et al., 2008).

Behavioral or *clinical scores* bring a rich description of the individuals involved in these studies. Importantly, they go beyond a diagnosis system for psychiatric patients (Insel and Cuthbert, 2015; Hyman, 2007), that can be overly strict and sometimes subjective. Indeed, the neuropathophysiology is often continuous, and many mental disorders are spectrum, understood as dimensional phenotypes rather than healthy-versus-sick categories. They may encompass several clinical subtypes of diseases that are only captured in detailed neuro-cognitive descriptions of the individuals (Fereshtehnejad et al., 2015; Insel et al., 2010). Importantly, while each behavioral variable gives a limited and noisy information, considering them jointly brings a more complete clinical picture, as diagnosis is usually based on combinations of these behavioral assessments.

Predictive models can extract neurophenotypes by mapping brain imaging data to individual traits. The corresponding imaging features typically predict either each cognitive assessment in isolation or a composite score that reflects these assessments. As a result, they are highly sensitive to the reliability of the score used, and probably under-exploit the rich clinical picture provided by the different assessments. To overcome these shortcomings, some studies use canonical correlation analysis (CCA) to relate multivariate imaging measurements to multiple scores: e.g. they capture co-variations between brain functional connectivity and a set of lifestyle, demographic, and clinical questionnaires, that are then mixed into one compound variable (Smith et al., 2015). On UK biobank's 5000-subjects cohort, Miller et al. (2016) found 9 basic modes of signal variation bridging brain images and traits, that mix information of very different nature. CCA relates multiple blocks of data (imaging features on one hand, behavioral scores on the other hand) through a latent factor model. However, it is by construction a global predictor on a combination of clinical score and does not aim at predicting well each individual score.

Here we are interested in good prediction of each score from neuroimaging features, at the single-subject level. We show that jointly learning multiple scores from multiple sources indeed gives better descriptions of subjects and mitigates the noise inherent to cognitive assessments. For this purpose, we use multi-output (a.k.a. multi-task) learning models (Caruana, 1997; Argyriou et al., 2008; Bzdok et al., 2015) to predict jointly clinical

scores (*outputs*) from neuroimaging data (*sources*). [1]

For population-imaging studies, resting-state functional MRI (rs-fMRI) is an interesting modality: it probes brain activity, and is a promising source of functional connectivity biomarkers. It is also easy to acquire and to compare across subjects, including on diminished subjects. Brain connectivity at rest is well suited to study pathologies (Greicius, 2008) as it is suitable for diminished patients and is impacted by neurodegenerative (Wang et al., 2006), or neuropsychiatric disorders (Craddock et al., 2009; Castellanos et al., 2013; Abraham et al., 2016). However, predicting clinical status of psychiatric or neurological patients from these data is challenging due to the low signal-to-noise ratio of the blood-oxygen-level dependent (BOLD) signal observed at rest.

In this paper, we show that fitting a richer clinical assessment of subjects based on multiple sources enhances rs-fMRI sensitivity and captures better neurophenotypes. The ensuing model exploits the correlation between multiple scores measuring subject health or behavioral outcomes, and combines information accumulated from several functional-connectivity maps. Technically, this approach combines multiple sources –connectivity maps or imaging modalities– for the joint prediction of multiple outputs –*i.e.* clinical scores. It is thus a *multi-modal* and *multi-output* model. We show that this approach improves the prediction of *each* output compared to single-output models and low-rank models like CCA. We provide an extensive validation on three different open datasets that characterize neurodegenerative (Alzheimer's Disease) and psychiatric disorders (schizophrenia). Going one step further, we evaluate this approach across datasets: predictive models –neurophenotypes– are learned on a cohort, but applied on another one with the same clinical scores and imaging modality (rs-fMRI). Our results show that modeling jointly multiple scores increases accuracy in the context of cross-cohort heterogeneity. Finally, we extend this framework by *stacking* heterogeneous sources –imaging, non-imaging– in multi-output models, and show the benefits of *stacking* compared to other multi-source combination strategies, such as multiple kernel learning or feature concatenation.

2. Materials and methods

2.1. Related works

In predictive settings, multi-task learning methods (Caruana, 1997; Argyriou et al., 2008) have been successfully applied to predict multi-output health outcomes from neuroimaging features. Notably, they have been used to fit individual memory performance scores (Wang et al., 2011), or to learn disease subcategories (Wang et al., 2015) and progression (Wang et al., 2012), or to predict a task from fMRI contrast maps of multiple subjects (Marquand et al., 2014), or to handle missing data (Yuan et al., 2012).

[1] We refer in this paper to multi-task models as *multi-output models*, to avoid ambiguity of interpretation of the term *task*.

Neuroimaging studies often use low-rank multi-output models like CCA (Hotelling, 1936) and partial least squares (PLS) (Krishnan et al., 2011) to link imaging based features to other blocks of data: to cross-predict fMRI and EEG (Deligianni et al., 2014), to explain genetic outcomes (Floch et al., 2012), behavioral and clinical scores (Monteiro et al., 2016; Miller et al., 2016; Smith et al., 2015), or a different imaging modality (Avants et al., 2010; Sui et al., 2012). Reduced rank regression (Vounou et al., 2010; Izenman, 1975) is a related multi-output linear-regression. Unlike CCA, it does not discard explained variance during model fitting.

Our goal is to predict several outputs from many – possibly heterogeneous – sources (*multi-output* and *multi-source*). Zhang et al. (2012) have tackled multiple source/output prediction as joint feature selection with a multi-task / multi-modal model. More in detail, they use multi-output models to perform a feature selection across scores on each modality. Each score is then predicted separately by combining multiple sources with a multiple kernel learning method (MKL) (Castro et al., 2014; Hinrichs et al., 2011). It is not a fully multi-output model since clinical scores are not learned jointly when combining different sources.

The model that we propose is designed similarly as in Zhang et al. (2012) but is completely multi-output, as it learns to jointly predict multiple scores from each source. Instead of using MKL to combine all sources, we use a versatile prediction stacking model as in Liem et al. (2016); Rahim et al. (2016). It combines predictions from multiple sources in another multi-output learning scheme.

2.2. Overview of the methodological framework

Multi-output prediction from multiple sources. We want to improve the prediction of health outcomes from multiple connectivity sources computed from rs-fMRI. Figure 1 depicts the workflow of the proposed model. It is a multi-output learning model built with a two-layer architecture. The first step performs a supervised dimension reduction with a multi-output classification or regression on each source separately. This yields a multi-output model for each source. Then, predictions corresponding to all sources are stacked score-wise: each score is associated with the first level prediction, which forms a subjects $\times$ sources matrix. The second level learns a non-linear combination of all these predictions in a common multi-output model, and returns the final prediction of the clinical scores, based on all data sources.

Seed-based connectivity maps. In our rs-fMRI experiments, sources are seed-based connectivity maps associated with different regions of interests (ROIs) from a brain atlas. A seed-based connectivity map is composed of correlations between time-series of a given seed ROI and voxels of the whole brain. To avoid relying on a single ROI, we use a set of ROIs extracted from a brain atlas $\mathcal{A}$. Each subject is represented by the resulting set of connectivity maps from which the predictive model is built.

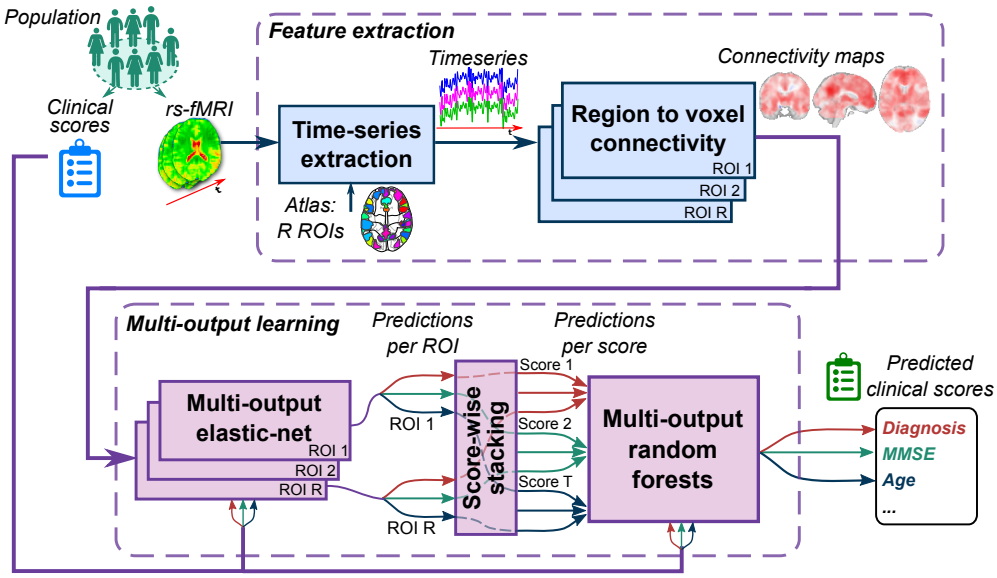


Figure 1: **Overview of the proposed method.** Each subject from a population is characterized by imaging data (e.g rs-fMRI) and clinical/behavioral data (e.g. phenotypes, scores). First, time-series are extracted from rs-fMRI. They are used to compute connectivity maps between ROIs and all brain voxels. Then, multi-output models are learned jointly across scores, yielding one multi-output model per source (in the present case, per ROI). Predictions from all sources are stacked score-wise (each score has a subjects × sources matrix). Finally, a multi-output random forest model selects the relevant sources (ROIs) that predict all clinical scores.

2.3. Stacked multi-output model

Table 1 summarizes the notations that we use in the description of the model. The principle of the proposed model is to first learn for each feature set (*source* $\mathbf{X}_i$) to predict jointly multiple clinical scores (*outputs* $\mathbf{Y}$). Then, a decision is learned to determine the final prediction of the clinical scores by combining all available predictions ($\mathbf{S}$). These two steps are performed as follows:

1) Sparse multi-output model for each input. Such models are well suited for high-dimensional connectivity features (around 10^5 brain voxels). They assume that the relevant information is located in a limited number of features, hence they reduce the dimension to a subset of non-zero coefficients. Linear models estimate a linear combination of voxel features, yielding a coefficient vector for each source. An important benefit of such models is that the coefficient vector can be interpreted as the discriminative map for the connectivity-based classification/regression.

We use a *multi-task elastic-net* regression (Argyriou et al., 2008). It can enforce more or less sparsity for a given amount of regularization. The multi-task specificity is that the sparsity is common to the coefficients for the different outputs. For the i^{th} source with a input matrix $\mathbf{X}_i \in \mathbb{R}^{N\times V}$ and a output matrix $\mathbf{Y} \in \mathbb{R}^{N\times T}$ (for N subjects, V features, and T clinical scores), the coefficient matrix $\widehat{\mathbf{W}}_i$ is estimated as:

$$\widehat{\mathbf{W}}_i = \underset{\mathbf{W}_i\in\mathbb{R}^{V\times T}}{\operatorname{argmin}} \frac{1}{2N}\|\mathbf{Y}-\mathbf{X}_i\mathbf{W}_i\|_F^2 + \alpha\left(\lambda\|\mathbf{W}_i\|_{2,1} + \frac{1-\lambda}{2}\|\mathbf{W}_i\|_F^2\right), \tag{1}$$

where $\alpha > 0$ controls the overall regularization parameter and $\lambda \in [0, 1]$ the sparsity of the estimate. $\|\mathbf{W}\|_{2,1}$ is the $\ell_{2,1}$ mixed norm of $\mathbf{W}$ (Kowalski, 2009) which is the key ingredient of group sparsity as in the group lasso (Yuan and Lin, 2006). The estimated weights related to the i^{th} source form a matrix $\widehat{\mathbf{W}}_i = [\widehat{\mathbf{w}}_{i,1}, \widehat{\mathbf{w}}_{i,2}, ..., \widehat{\mathbf{w}}_{i,T}]$, where each vector $\widehat{\mathbf{w}}_{i,j}$ represents the predictive coefficients for the j^{th} output. The $\ell_{2,1}$ sparsity prior promotes zero weights jointly across the coefficients related to the various outputs. In other words, coefficients of the different vectors $\widehat{\mathbf{w}}_{i,\cdot}$ are zero when they are not relevant across all outputs.

Table 1: Notations used in the method description.

N	number of subjects of the dataset.
V	number of voxels of the entire brain.
T	number of outputs (clinical scores).
R	number of sources.
$\mathcal{A}$	brain atlas: set of R brain regions a_i, such that: $\mathcal{A} = \{a_i\}$ with $1 \le i \le R$.
$\mathbf{X}_i$	Connectivity maps (sources) of all subjects according to the i^{th} ROI, its dimension is $N \times V$.
$\mathbf{y}_j$	vector of clinical scores associated with the j^{th} score, its dimension is N.
$\mathbf{Y}$	matrix of clinical scores, its dimension is $N \times T$.
$\mathbf{w}_{i,j}$	coefficient vector of the linear model associated with the i^{th} source and the j^{th} output. Its dimension is V
$\mathbf{W}_i$	coefficient matrix of all linear models associated with the i^{th} source and all outputs. Its dimension is $V \times T$.

2) Multi-output predictions stacking. First step predictions across different sources are often complementary and combining them efficiently increases prediction accuracy. Rather than averaging them or performing a majority voting, we propose to use another predictive model to optimally combine different per-source predictions. First, we define a multi-output prediction matrix according to a source i as the unthresholded predictions outputs $\mathbf{X}_i\widehat{\mathbf{W}}_i$ from (1). Then, all sources are stacked score-wise yielding a 3D tensor $\mathbf{S}$ of dimension $T \times N \times R$. For the j^{th} clinical score, the stacked matrix $\mathbf{S}_j \in \mathbb{R}^{N\times R}$ is:

$$\mathbf{S}_j = \left[\, \mathbf{X}_1\,\widehat{\mathbf{w}}_{1,j}, \ldots, \mathbf{X}_i\,\widehat{\mathbf{w}}_{i,j}, \ldots, \mathbf{X}_R\,\widehat{\mathbf{w}}_{R,j} \,\right]. \tag{2}$$

We use *random forests* (RF) to learn a mapping from stacked sources to multiple outputs. Such models make decisions by combining non-linearly a small number of sources and yield an importance index to quantify the contribution of each source to the final prediction. Regression random forests (Breiman, 2001) are an ensemble of regression trees, where a tree provides regression values (clinical scores) from input data (predictions according to multiple sources). Here we need a multi-output predictor since we have multiple inputs (stacked predictions $\mathbf{S}_j$) associated with multiple outputs (clinical scores $\mathbf{y}_j$): $\Big([\mathbf{S}_1, ..., \mathbf{S}_j, ..., \mathbf{S}_T], [\mathbf{y}_1, ..., \mathbf{y}_j, ..., \mathbf{y}_T]\Big)$. For this, decision trees are constructed by finding nodes that split data according to a feature value minimizing an impurity criterion Γ. In multi-output random forests, the impurity criterion is the sum of the impurities of each output:

$$\Gamma\,(\,\text{node}, \text{split}\,) = \sum_{j=1}^{T} \Gamma_j\,(\,\text{node}, \text{split}\,). \qquad (3)$$

The impurity criterion Γ_j specific to the j^{th} output is the squared error between outputs when choosing a split on a node:

$$\Gamma_j\,(\text{node}, \text{split}) = \sum_{k \in \text{node}} \Big(\mathbf{y}_j(k) - \text{mean}(\mathbf{y}_j \,|\, \text{node}, \text{split})\Big)^2, \qquad (4)$$

where the mean of $\mathbf{Y}_j$ is taken on the samples selected according to some optimal split at a given node in $\mathbf{S}_j$. The difference with respect to classical random forests is that we determine jointly the splits across all outputs by taking the sum of all the impurities Γ_j as the global impurity criterion (Segal and Xiao, 2011), as depicted in Figure 2. Random forests are then built by bootstrapping on training data and averaging the resulting tree-based predictions. This yields more stable predictions compared to single estimators. Finally, continuous random-forest outputs that represent binary classes are discretized, where probabilities higher and lower than 0.5 are positive and negative labels, respectively.

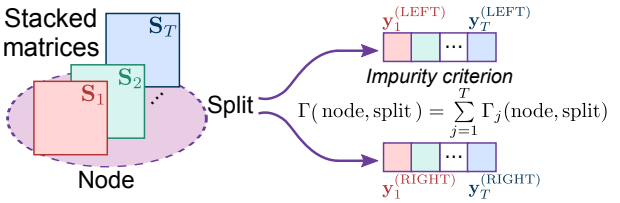


Figure 2: **Multi-output random forests principle.** Inputs are stacked predictions for each clinical score j. Nodes (subsets from stacking matrices $\mathbf{S}$) are divided according to an impurity criterion over all outputs to be predicted.

3. Experiments: learning multiple neuropsychiatric profiles from resting-state data

We now demonstrate empirically that jointly fitting multiple clinical, neurological, and neuropsychiatric scores yields better predictions than separate predictions. For this, we perform several experiments over three publicly available datasets with rs-fMRI that study different clinical questions, namely schizophrenia, post-traumatic stress disorders and Alzheimer's Disease.

Table 2: **Demographic characteristics of the datasets used in the experiments.** Clinical scores are detailed in table 3.

	COBRE	**ADNI**	**ADNIDOD**
# subjects	160	211	127
Age	38.2 ± 12.7	72.6 ± 6.8	68.7 ± 4.4
# scores	13	9	5
Groups	Schizophrenia (60) Control (72)	AD (57) MCI (154)	

3.1. Datasets

Table 2 summarizes the datasets used in our experiments. Table 3 lists clinical scores' acronyms. Details on Alzheimer's Disease and schizophrenia related scores can be found respectively in Petersen et al. (2010) and Calhoun et al. (2012).

– *COBRE.* The COBRE dataset characterizes the brain in schizophrenia based on rs-fMRI. It is provided from the Center for Biomedical Research Excellence (`cobre.mrn.org`). It comprises anatomical and functional MRI data from 72 patients with schizophrenia and 75 healthy controls. Diagnostic information and clinical scores were collected using the Structured Clinical Interview used for DSM Disorders (SCID).

– *ADNI.* The Alzheimer's Disease Neuroimaging Initiative (`www.adni-info.org`) database is a multi-modal study of brain aging and neuro-degenerative diseases. Our goal is to study conversion of MCI subjects to AD. We consider a proxy by estimating binary classification models to discriminate AD against MCI subjects. We use 211 subjects with one to five fMRI scans, resulting in a total of 694 fMRI scans. We also include hippocampus features extracted from anatomical MRI and biomarker measurements extracted from cerebrospinal fluid (CSF).

– *ADNIDOD.* The US Department of Defense (DoD) Study of Brain Aging in Vietnam War Veterans (`www.adni-info.org/DOD.html`) gathers imaging and non-imaging measures from persons who had post-traumatic stress disorders and / or traumatic brain injuries. The study investigates potential links between brain trauma and neuro-degenerative diseases. ADNIDOD uses the same fMRI acquisition protocol (same scanner type) and the same clinical and cognitive examination protocols as ADNI. We use rs-fMRI scans at baseline from 127 subjects, and 5 clinical scores common with ADNI. We use this dataset to validate predictive models learned on ADNI, but not to learn new ones.

3.2. Data preprocessing

For all datasets, we apply standard fMRI preprocessing: discarding first 3 frames from each scan, motion correction, fMRI coregistration to T1-MRI, normalization to MNI template using SPM12, (`www.fil.ion.ucl.ac.uk/spm/`), followed by spatial smoothing (6mm FWHM). Temporal preprocessing includes linear detrending and band filtering (0.01 – 0.1Hz). Then

Table 3: Clinical scores in COBRE, ADNI, and ADNIDOD datasets.

Dataset	Score	Description
COBRE	Age	Subject age at the acquisition date.
	BACS	Brief Assessment of Cognition in Schizophrenia
	MD ProcSpeed	MatricsDomain Processing Speed Index
	TMT A	Trail Making Test Assessment
	WAIS Coding	Wechsler Adult Intelligence Scale on Digital Symbol Coding
	WAIS PSI	Wechsler Adult Intelligence Scale Processing Speed Index
	WAIS SymSearch	Wechsler Adult Intelligence Scale on Symbol Search
	WASI Sim	Wechsler Abbreviated Scale Intelligence on Similarities
	WASI verb	Wechsler Abbreviated Scale Intelligence on Vocabulary
	Diagnosis	Schizophrenia or Controls
ADNI, ADNIDOD	Age	Subject age at the acquisition date.
	CDR	Clinical dementia rating. Possible values: $\{0, 0.5, 1, 2, 3\}$
	ADAS	Alzheimer's disease assessment scale. (ADNI only)
	MMSE	Mini-mental state examination. Possible values: $[0, 30]$
	FAQ	Functional activities questionnaire.
	NPIQ	Neuropsychiatric inventory questionnaire.
	GDS	Geriatric depression scale.
	LDEL	Logical memory delayed.
	NSS	Neuropsychiatric summary scores. (ADNI only)
	Diagnosis	AD or MCI diagnosis group. (ADNI only)

a matrix of confounds is built for each subject. It contains age, white matter, 12 motion components – 6 motion parameters and their first order derivatives – and 6 noise components from CompCor (Behzadi et al., 2007). Confounds are removed by subtracting the time-series projected into an orthonormal basis of the confounds, using Nilearn v0.2.6 (Abraham et al., 2014). All rs-fMRI scans are quality-checked. In addition to the visual inspection, we apply some simple rules: Scans are excluded when motion is higher than 2mm or when global signal variation is higher than 5%. Excluded subjects in ADNI and COBRE are listed in the supplementary material.

Anatomical MRI and CSF features are collected from processed and quality-checked data uploaded from ADNI database. For the CSF, three biomarkers measurements ($A\beta_{1\text{-}42}$, t-tau, p-tau$_{181}$) are extracted from an analysis done at the University of Pennsylvania (Shaw et al., 2011). Based on anatomical MRI, we select sixteen volumetric features of segmented hippocampus extracted with `FreeSurfer` software (Dale et al., 1999) at the university of California in San Francisco. Beside the diagnosis –Alzheimer's Disease (AD), Mild Cognitive Impairment (MCI), schizophrenia, or controls– we predict Neuropsychiatric outputs form of continuous scores. We apply a Box-Cox transformation on continuous clinical scores. It yields a distribution of the outputs with more Gaussian-like characteristics and more stable variances. The optimal transformation parameters are estimated such that they maximize the profile log-likelihood function (Box and Cox, 1964).

3.3. Experimental settings

We present three experiments studying the proposed multi-output framework on rs-fMRI: in a single dataset, across datasets, and in a multi-modal setting. They assess:

(i) The multi-output approach compared to single-output on a single cohort. This experiment is carried on COBRE, to predict schizophrenia, and on ADNI, to predict AD profiles. Models are evaluated with a cross-validation procedure as recommended in Varoquaux et al. (2016): The data are split into stratified train/test sets at the subject level, to avoid fitting and testing on data from the same subject. Splits are randomized over 100 runs. Test sets represent 25% of the whole dataset. Single-output prediction relies on stacking elastic-net predictions from each source by a single-output random-forest.

(ii) The multi-output approach across datasets. This is done with a cross-dataset learning (train on ADNI, test on ADNIDOD); in the cross cohort validation, the randomized splits are done over 90% of each dataset, as depicted in Figure 3.

(iii) Stacking compared to other source combination strategies. We show the benefits of using prediction stacking for classification by comparing it to classical model aggregation strategies such as majority voting or prediction averaging.

(iv) Stacked model versus other multi-output analysis approaches. This experiment explores whether using the proposed stacked multi-output model is more accurate than other existing multi-output methods to predict clinical scores. For this, we compare the stacked multi-output model to CCA and reduced rank regression (RRR) to predict clinical scores. CCA finds linear combinations to optimize correlations between $\mathbf{X}$ and $\mathbf{Y}$. RRR is a multi-output linear regression that imposes a low-rank constraint on the coefficient matrix (Izenman, 1975). This model seeks common latent factors across the different outputs to improve regression accuracy. CCA is not a direct predictive model, but predictions of $\mathbf{Y}$ can be derived from learned components. They can be used then on test samples in order to measure the fidelity of the CCA on unseen data. We assess the prediction capacity of low-rank linear models compared to our stacked multi-output model. We use the concatenated connectivity maps as features for the CCA and RRR. The number of components of CCA and RRR is set to 2 by applying a nested cross-validation. Accuracies of these two models are compared to the stacked multi-output model by shuffling train and test splits over 100 runs.

(iv) Stacked multi-output model for multi-modality fusion. In this experiment, AD is characterized from heterogeneous modalities (anatomical, functional, biofluid) in ADNI. For combining the modalities, we compare our multi-output approach –based on prediction stacking– to other schemes: simple feature concatenation or multiple kernel learning.

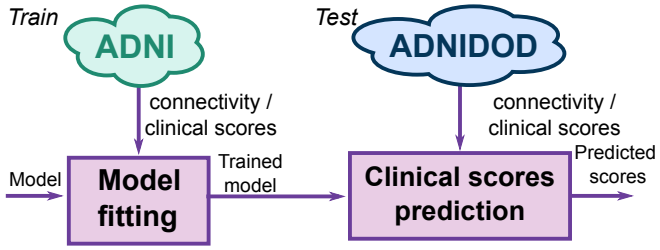


Figure 3: **Cross-datasets validation scheme.** Single-output and stacked multi-output models are trained on all samples / features / scores of ADNI. Model accuracy is assessed on clinical scores that are common with ADNIDOD. Confidence intervals are obtained through randomized subsets of ADNIDOD.

– *Assessing classification performance.* We assess classification performance with the area under the receiver operating characteristic curve (AUC). It quantifies the sensitivity (true positive rates) and the specificity (false positive rates).

– *Assessing regression performance.* We measure accuracy of clinical score regressions on each test fold with the cross-validated determination coefficient[2]:

$$\Delta_{\text{cv}} = 1 - \frac{\sum_{k=1}^{|\text{test}|} (y_k - \widehat{y}_k)^2}{\sum_{k=1}^{|\text{test}|} (y_k - \bar{y}_{\text{test}})^2} \qquad (5)$$

where $y_k, \widehat{y}_k, \bar{y}_{\text{test}}$ are respectively the true score, the predicted score, and the mean score on the test fold.

In all experiments, elastic-net hyper-parameters (α, λ) are set by a nested 4-fold cross-validation. They are estimated inside each training set for each source. For the multi-output random forest, we set the number of trees and the maximum depth of a tree to commonly-used values, 50 and 10 respectively.

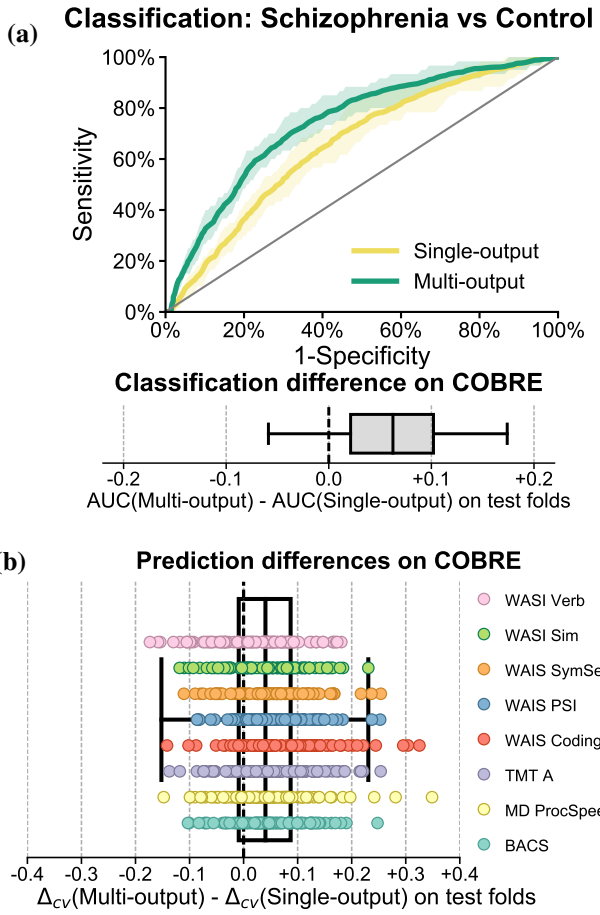
The significance of the value of Δ_{cv} for each model and each score is assessed by permutation tests with 10,000 permutations. Single and multi-output models are compared in each experiment through their respective Δ_{cv} on the same folds. Finally, experiments are implemented in `Python` using Scikit-Learn v0.17 (Pedregosa et al., 2011).

4. Results: performance factors in multi-output learning

4.1. Multi-output learning predicts better each score

We compare multi-output and single-output approaches for the prediction of clinical scores on COBRE and ADNI.

[2]Many studies use Pearson's correlation to evaluate regression accuracy. Correlations measure the linear accordance between true and predicted values, but they discard scaling and offsets. Using Δ_{cv} brings a better assessment of regression quality –as compared to a mean-constant model. Also called r-squared metric, it is a relative distance that can be used to compare different models. Unlike correlations, it is sensitive to absolute differences between the prediction values and the true values, which is essential in the perspective of individualized predictions.



(c)

Clinical scores	Single-output Δ_{cv}	Multi-output Δ_{cv}
WASI Verb	-0.04±0.06	-0.01±0.09
WASI Sim	0.08±0.07	-0.01±0.08
WAIS SymSearch	0.09±0.07	**0.18±0.09**
WAIS PSI	0.09±0.06	**0.18±0.08**
WAIS Coding	0.08±0.07	**0.19±0.10**
TMT A	-0.04±0.07	0.05±0.09
MD ProcSpeed	0.03±0.07	0.09±0.10
BACS	**0.16±0.07**	**0.24±0.08**

	Single-output AUC	Multi-output AUC
Classification	**0.65 ± 0.11**	**0.74 ± 0.08**

Figure 4: **Comparison between single / multi-output learning on COBRE.** Classification results **(a)** show better sensitivity and specificity for the multi-output model. Regression results **(b)** show improvements with the multi-output model. **(c)** shows Δ_{cv} means and standard deviations over 100 shuffled test folds. Bold indicates statistically significant accuracy ($p < .001$ with a permutation test).

– *Predicting schizophrenia characteristics.* Figure 4 compares accuracies of multi-output and single-output models on the COBRE dataset. The ROC curves (Figure 4a) show that, compared to the single-output model, the stacked multi-output model has higher sensitivity to predict schizophrenia and also higher specificity against false positives. Additionally, the AUC has less variance over randomized folds for the stacked multi-

output model. A gain of .1 in AUC, or 15% sensitivity at 50% specificity is substantial. Learning the diagnosis jointly with related quantitative clinical scores yields more accurate and more stable prediction of schizophrenia status. Overall, these results confirm previous successes predicting schizophrenia from rs-fMRI (Savio and Graña, 2015; Rashid et al., 2016).

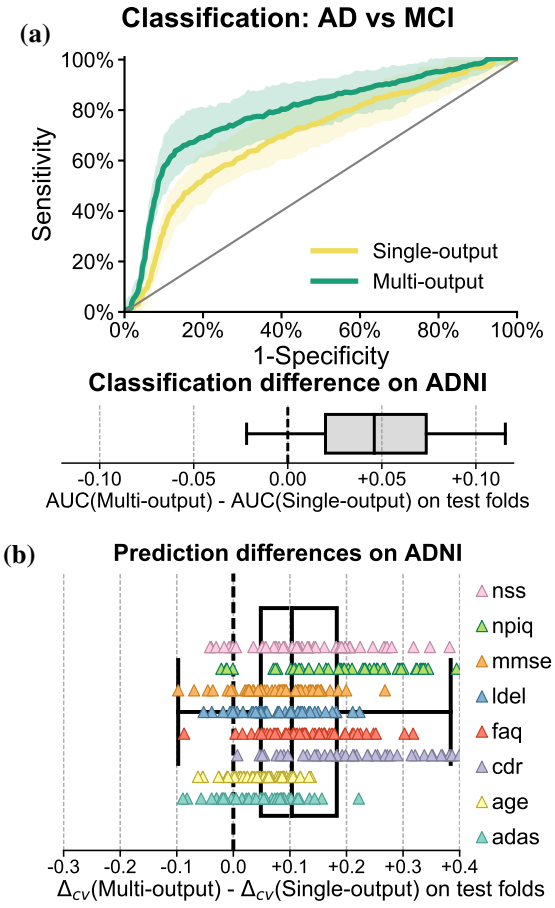
Figure 4b shows that, for some cognitive scores, prediction from functional connectivity achieves only a limited accuracy: some cross-validation determination coefficients Δ_{cv} are negative (*WASI verb, WASI Sim*), meaning that even the mean of the test values is hard to predict. Several hypotheses can be considered. Besides the low SNR captured in the BOLD signal and the low number of subjects, some cognitive scores may not be reflected in functional connectivity. However, the stacked multi-output model improves the overall Δ_{cv} for all scores, and notably the Brief Assessment Cognition in Schizophrenia (*BACS*), that is considered as a reliable and reference evaluation metric of schizophrenia (Keefe, 2004). Differences between Δ_{cv} of the two models are positive in average, meaning that the stacked multi-output model is improving the overall accuracy.

– Predicting AD characteristics. Comparing single and multi output models on the ADNI dataset show results similar to those on the COBRE dataset. This confirms the positive impact of using a stacked multi-output model to characterize neuropsychiatric phenotypes from resting-state fMRI features. In Figure 5a, cross-validated classification shows better AD diagnosis prediction with the stacked multi-output model.

For AD-related clinical scores regression, Δ_{cv} differences between multi-output and single output models are in average around 0.1. Average Δ_{cv} over shuffled train and test splits shows that for some clinical scores prediction is not better than chance. This holds for *npiq* where the test set mean is poorly estimated ($\Delta_{cv} < 0$). This can be explained by the fact that this score does not yield a very sensitive characterization of dementia, nor does it differentiate between clinical groups (Lai, 2014). Also, the clinical dementia rate (*CDR*) has only 5 possible values, which makes regression approaches less accurate. However, predictions of most of the clinical scores (*MMSE*, *ADAS* ...) are statistically significant, and the multi-output model improves significantly the overall accuracy in all cases.

4.2. Multi-output stacking reduces variability across datasets

We study to what extent a model learned from a study generalizes to a different study with the same measures –rs-fMRI as input, clinical scores as outputs– but on different cohorts. Figure 6 shows that multi-output improves overall accuracy markedly for this cross-dataset prediction. These results also show that predicting across datasets is hard: accuracies are lower than intra dataset. This is due to the variability between studies (sites, scanner, ...), coupled to the low SNR captured by rs-fMRI. As mentioned in (Woo et al., 2017), predicting across sites and datasets is still an open-issue, and few studies attempted to tackle this question. However, capturing jointly several outputs in the predictive model limits the overfit of study idiosyncrasies that is detrimental to prediction accuracy.



(c)

Clinical scores	Single-output Δ_{cv}	Multi-output Δ_{cv}
nss	0.04±0.14	**0.20±0.13**
npiq	-0.25±0.16	-0.01±0.09
mmse	0.05±0.10	**0.14±0.09**
ldel	**0.12±0.10**	**0.18±0.09**
faq	0.01±0.15	**0.14±0.11**
cdr	-0.23±0.20	0.06±0.11
age	**0.14±0.14**	**0.20±0.10**
adas	**0.13±0.12**	**0.16±0.08**

	Single-output AUC	Multi-output AUC
Classification	**0.72 ± 0.05**	**0.76 ± 0.04**

Figure 5: **Comparison between single / multi-output learning on ADNI. (a)** ROC curves and their confidence interval for AD/MCI classification; the stacked multi-output model outperforms single-output models. **(b)** Absolute differences between multi-output and single output clinical score regression (Δ_{cv}). Each dot represents a difference on a test fold. A median gain of 0.1 is obtained when using stacked multi-output model. **(c)** shows Δ_{cv} means and standard deviations over 100 shuffled test folds. Bold values indicate statistically significant accuracy ($p < .001$ with a permutation test).

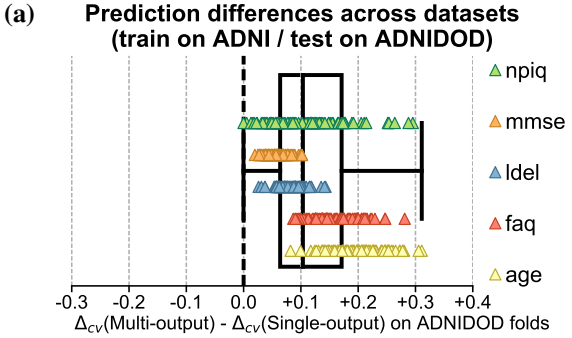
4.3. Stacking outperforms other source combination strategies

We show in this experiment the benefits of using random forests to make final predictions from each source predictions. We compare random forests based stacking against model averaging and majority voting to predict the diagnosis for CO-

BRE and ADNI datasets. Figure 8 summarizes multi-output classification accuracies according to each prediction combination strategy, using the same experiment design as presented in the manuscript (100 randomizations over train-test splits). We observe that stacking the predictions consistently outperforms majority voting and averaging. Differences with respect to the mean are increased on average by +4% and +2% for COBRE and ADNI, respectively. We also observe that the accuracies are more stable with the stacking, in particular for schizophrenia prediction.

4.4. *Stacked multi-output models predict better than low-rank multi-output models*

We compare regression Δ_{cv} on COBRE and ADNI test folds with RRR, CCA, and the stacked multi-output model. Figure 7 shows that stacked multi-output model has the best overall prediction accuracy compared to CCA and RRR. CCA accuracy on test folds for each clinical score are not significant since most of average Δ_{cv} are below zero. This is not surprising, since CCA is not a model optimized for prediction. It is a latent factor analysis method that aims at detecting linear relationships between imaging features and clinical scores. RRR's Δ_{cv} are lower than those of the stacked multi-output method. Also, we observe that discrepancies between the accuracies of the clinical scores are consistent across the three methods. For example, *BACS*, *WAIS Coding* is better predicted than *WASI Sim* and *WASI verb* with all methods. This suggests that these scores better characterize the schizophrenia phenotype. Similar trends are observed when comparing the multi-output model to CCA and RRR on the ADNI dataset. To summarize, these results show that the stacked multi-output model outperforms CCA and RRR to predict jointly multiple clinical and behavioral scores.



(b)

	Single Δ_{cv}	Multi Δ_{cv}
npiq	-0.33±0.16	-0.13±0.06
mmse	-0.01±0.12	**0.08±0.04**
ldel	-0.01±0.05	**0.10±0.04**
faq	-0.21±0.08	0.05±0.04
age	-0.25±0.11	0.01±0.06

Figure 6: **Cross-dataset validation of multi-output stacking.** **(a)** Multi-output stacking yields better accuracy than single output models in all test folds. **(b)** shows means and standard deviations of Δ_{cv} over 100 shuffled test folds. Bold indicates statistically significant accuracies with a permutation test.

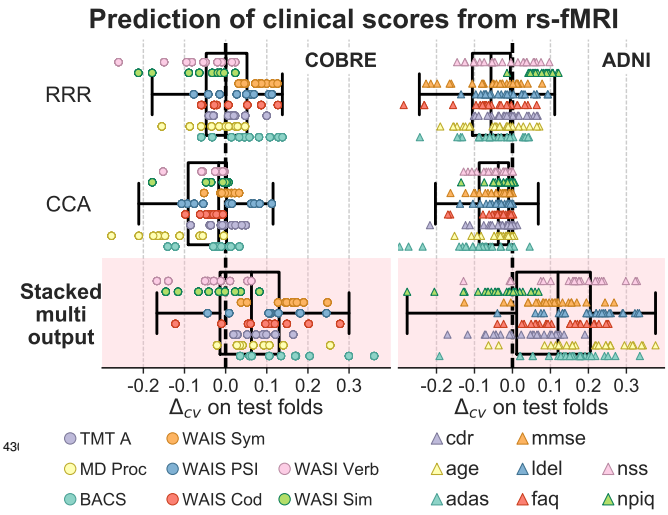
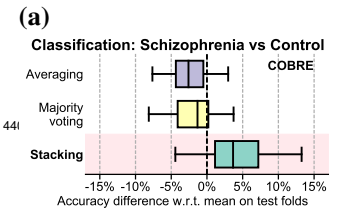


Figure 7: **Comparing multi-output stacking to Canonical Correlation Analysis CCA and Reduced Ranked Regression (RRR) on COBRE and ADNI.** Stacked multi-output models predict better than other multi-output analysis methods. Results are consistent across datasets.



(c)

Dataset	averaging	majority voting	stacking
COBRE	66.9% ± 6.2%	67.7% ± 6.5%	**73.6% ± 3.0%**
ADNI	73.7% ± 3.8%	74.2% ± 3.7%	**76.1% ± 3.7%**

Figure 8: **Comparing different multi-output source combination strategies in COBRE and ADNI.** – **(a, b)** Distributions of accuracy differences with respect to subject-wise means shows that stacking connectivity sources outperforms model averaging and majority voting. **(c)** Accuracies of source combination strategies: Mean and standard deviations of connectivity-based prediction accuracy according to model averaging, majority voting and stacking.

4.5. *Multi-output stacking improves multi-modal prediction*

This last experiment demonstrates using the proposed multi-output stacking approach on multi-modal data. On the ADNI dataset, predicting four clinical scores, we compare the multi-output approach with single output prediction, first with each modality separately (rs-fMRI, hippocampus, CSF), then by combining modalities with a simple feature concatenation or a multiple kernel learning (MKL) method that is considered as a standard for multi-modal/dimensional learning (Schrouff et al., 2016; Gönen et al., 2011). Figure 9 shows that predicting several clinical scores improves accuracies over learning each score separately. CSF and hippocampus-only based predictions remain similar. This is not surprising, as the feature space is small (3 and 16 respectively), which limits the effect of the multi-output feature selection.

Combining all modalities increases prediction accuracies, in particular when the stacking approach is used. On average, the

stacking approach leads to higher accuracies than MKL and feature concatenation.

5. Discussion and conclusion

We have introduced a stacked multi-output learning framework that leverages many input sources to jointly predict several clinical scores. It is a competitive alternative to classical multi-output analysis models like CCA or RRR. Systematic comparisons show that the stacked multi-output models predict more accurately neuropsychiatric profiles or diagnostic status from neuroimaging measures, in particular functional-connectivity maps.

Multi-output models improve the prediction of the different clinical scores. This is because they exploit similar and complementary quantifications of patient health outcomes, hence overcome the limitations of both the modality and the behavioral assessments (small dataset, noisy data). In both applications that we studied –schizophrenia and Alzheimer's disease– clinical-assessment scores complement the disease / healthy evaluation. Each score is individually noisy and difficult to predict. However, our experiments show that capturing them jointly in a multi-output approach not only improves the prediction of these scores but also yields a net increase in patient / control classification.

Additionally, stacked multi-output models bring sizable accuracy gains when predicting to a new cohort studying brain disorders, here generalizing from the ADNI database to the ADNIDOD cohort. These results show that capturing jointly several clinical targets has a crucial impact in the face of data heterogeneity.

The proposed framework can include multi-modal information. Multi-output feature selection and prediction stacking together make the method more accurate when combining anatomical, functional, and biofluid biomarkers. The experimental results show that multi-output stacking is as accurate as state-of-the-art multi-modal approaches such as multiple kernel learning. From a practical point of view, it is beneficial because it requires less parameter tuning and the prediction stacking step can easily incorporate any predictor or signal capturing a new source of information.

Future work calls for leveraging the proposed approach to build cross-study predictive models from the growing large multi-modal databases. Resources like UK biobank can provide novel hypotheses and biomarkers that can be tested in targeted clinical investigations. These rich cohorts typically provide multiple clinical and cognitive scores. Our cross-dataset results suggest that, analyzed jointly, these scores will yield biomarkers that generalize better to new cohorts and provide new insights linking neuropathologies to the brain.

Acknowledgements

This work is supported by the Lidex PIM project funded by the IDEX Paris-Saclay (ANR-11-IDEX-003-02) and the NiConnect project (ANR-11-BINF-0004_NiConnect).

References

Abi-Dargham, A., Horga, G., 2016. The search for imaging biomarkers in psychiatric disorders. Nature Medicine 22, 1248.

Abraham, A., Milham, M., Martino, A.D., Craddock, R.C., Samaras, D., Thirion, B., Varoquaux, G., 2016. Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example. NeuroImage .

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., Varoquaux, G., 2014. Machine learning for neuroimaging with scikit-learn. Frontiers in Neuroinformatics 8, 14.

Argyriou, A., Evgeniou, T., Pontil, M., 2008. Convex multi-task feature learning. Machine Learning 73, 243–272.

Avants, B.B., Cook, P.A., Ungar, L., Gee, J.C., Grossman, M., 2010. Dementia induces correlated reductions in white matter integrity and cortical thickness: A multivariate neuroimaging study with sparse canonical correlation analysis. NeuroImage 50, 1004–1016.

Behzadi, Y., Restom, K., Liau, J., Liu, T.T., 2007. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. NeuroImage 37, 90–101.

Box, G.E., Cox, D.R., 1964. An analysis of transformations. Journal of the Royal Statistical Society. Series B (Methodological) , 211–252.

Breiman, L., 2001. Random forests. Machine learning 45, 5–32.

Bzdok, D., Eickenberg, M., Grisel, O., Thirion, B., Varoquaux, G., 2015. Semi-supervised factored logistic regression for high-dimensional neuroimaging data, in: NIPS 2015, pp. 3348–3356.

Calhoun, V.D., Sui, J., Kiehl, K., Turner, J., Allen, E., Pearlson, G., 2012. Exploring the psychosis functional connectome: Aberrant intrinsic networks in schizophrenia and bipolar disorder. Frontiers in Psychiatry 2.

Caruana, R., 1997. Multitask Learning. Machine Learning 28, 41–75.

Castellanos, F.X., Martino, A.D., Craddock, R.C., Mehta, A.D., Milham, M.P., 2013. Clinical applications of the functional connectome. NeuroImage 80.

Castro, E., Gómez-Verdejo, V., Martínez-Ramón, M., Kiehl, K.A., Calhoun, V.D., 2014. A multiple kernel learning approach to perform classification of groups from complex-valued fMRI data analysis: Application to schizophrenia. NeuroImage 87, 1–17.

Craddock, R.C., Holtzheimer, P.E., Hu, X.P., Mayberg, H.S., 2009. Disease state prediction from resting state functional connectivity. Magnetic Resonance in Medicine 62, 1619–1628.

Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis. NeuroImage 9, 179–194.

Deligianni, F., Centeno, M., Carmichael, D.W., Clayden, J.D., 2014. Relating resting-state fMRI and EEG whole-brain connectomes across frequency bands. Frontiers in Neuroscience 8.

Di Martino, A., Yan, C.G., Li, Q., Denio, E., Castellanos, F.X., Alaerts, K., Anderson, J.S., Assaf, M., Bookheimer, S.Y., Dapretto, M., et al., 2014. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular psychiatry 19, 659.

Drysdale, A.T., Grosenick, L., et al., 2016. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. Nature Medicine .

Dubois, J., Adolphs, R., 2016. Building a science of individual differences from fmri. Trends in cognitive sciences 20, 425–443.

Elliott, P., Peakman, T.C., et al., 2008. The UK biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. International Journal of Epidemiology 37, 234–244.

Ellis, K.A., Bush, A.I., Darby, D., Fazio, D.D., Foster, J., Hudson, P., Lautenschlager, N.T., Lenzo, N., Martins, R.N., Maruff, P., Masters, C., Milner, A., Pike, K., Rowe, C., Savage, G., Szoeke, C., Taddei, K., Villemagne, V., Woodward, M., Ames, D., 2009. The australian imaging, biomarkers and lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of alzheimer's disease. International Psychogeriatrics 21, 672.

Fereshtehnejad, S.M., Romenets, S.R., Anang, J.B.M., Latreille, V., Gagnon, J.F., Postuma, R.B., 2015. New clinical subtypes of parkinson disease and their longitudinal progression. JAMA Neurology 72, 863.

Floch, E.L., Guillemot, V., et al., 2012. Significant correlation between a set of genetic polymorphisms and a functional brain network revealed by feature selection and sparse partial least squares. NeuroImage .

Gönen, M., Kandemir, M., Kaski, S., 2011. Multitask learning using regularized multiple kernel learning, in: Neural Information Processing, Springer. pp. 500–509.

Greicius, M., 2008. Resting-state functional connectivity in neuropsychiatric disorders. Current opinion in neurology 21, 424.

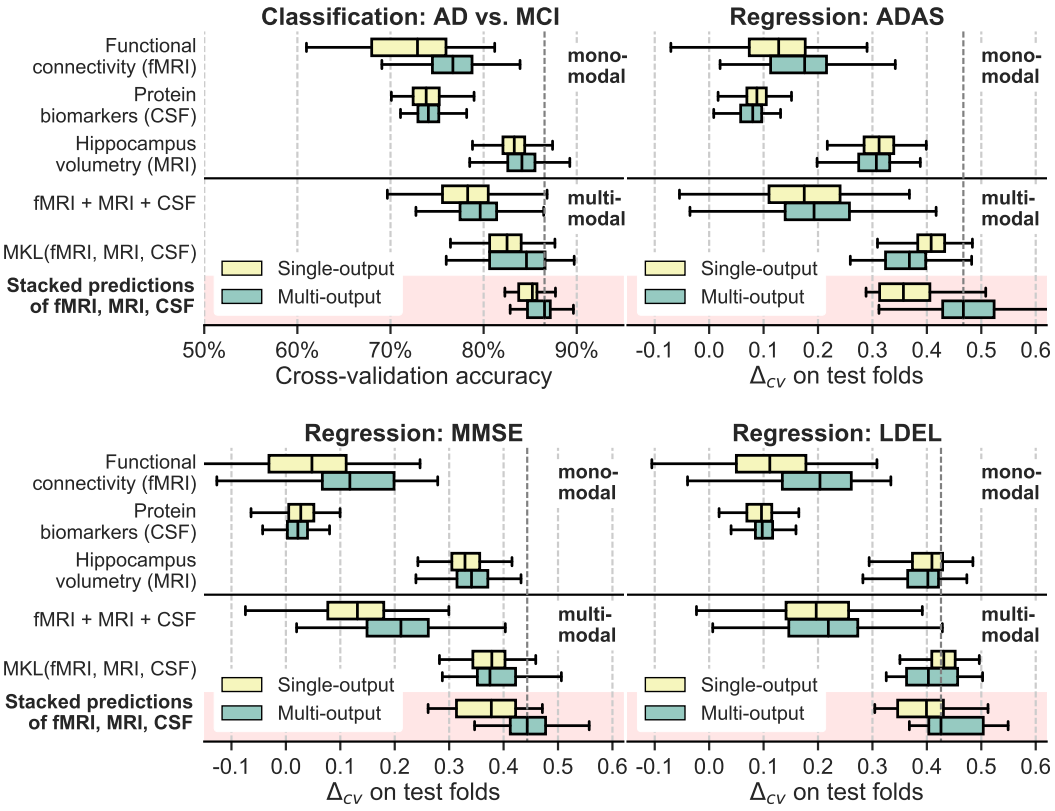


Figure 9: **Predicting clinical scores with multi-modal data on the ADNI dataset.** Extending stacked multi-output model to combine complementary modalities improves predictions over using a single modality. Prediction stacking is at least as accurate as MKL methods.

Hinrichs, C., Singh, V., Xu, G., Johnson, S.C., 2011. Predictive markers for AD in a multi-modality framework: An analysis of MCI progression in the ADNI population. NeuroImage 55, 574–589.

Hotelling, H., 1936. Relations between two sets of variates. Biometrika 28, 321–377.

Hyman, S.E., 2007. Can neuroscience be integrated into the DSM-v? Nature Reviews Neuroscience 8, 725–732.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D.S., Quinn, K., Sanislow, C., Wang, P., 2010. Research domain criteria (rdoc): toward a new classification framework for research on mental disorders. American Journal of Psychiatry 167, 748–751.

Insel, T.R., Cuthbert, B.N., 2015. Brain disorders? precisely. Science 348, 499–500.

Izenman, A.J., 1975. Reduced-rank regression for the multivariate linear model. Journal of multivariate analysis 5, 248–264.

Jack, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L Whitwell, J., Ward, C., et al., 2008. The alzheimer's disease neuroimaging initiative (adni): Mri methods. Journal of Magnetic Resonance Imaging 27, 685–691.

Keefe, R., 2004. The brief assessment of cognition in schizophrenia: reliability, sensitivity, and comparison with a standard neurocognitive battery. Schizophrenia Research 68, 283–297.

Kowalski, M., 2009. Sparse regression using mixed norms. Applied and Computational Harmonic Analysis 27, 303–324.

Krishnan, A., Williams, L.J., McIntosh, A.R., Abdi, H., 2011. Partial least squares (PLS) methods for neuroimaging: A tutorial and review. NeuroImage 56, 455–475.

Lai, C., 2014. The merits and problems of neuropsychiatric inventory as an assessment tool in people with dementia and other neurological disorders. CIA , 1051.

Liem, F., Varoquaux, G., Kynast, J., Beyer, F., Masouleh, S.K., Huntenburg, J.M., Lampe, L., Rahim, M., Abraham, A., Craddock, R.C., Riedel-Heller, S., Luck, T., Loeffler, M., Schroeter, M.L., Witte, A.V., Margulies, D.S., 2016. Predicting brain-age from multimodal imaging data captures cognitive impairment. NeuroImage .

Loeffler, M., Engel, C., et al., 2015. The life-adult-study: objectives and design of a population-based cohort study with 10,000 deeply phenotyped adults in germany. BMC public health 15, 1.

Marquand, A.F., Brammer, M., Williams, S.C., Doyle, O.M., 2014. Bayesian multi-task learning for decoding multi-subject neuroimaging data. NeuroImage 92, 298–311.

Miller, K.L., Alfaro-Almagro, F., et al., 2016. Multimodal population brain imaging in the UK biobank prospective epidemiological study. Nature Neuroscience .

Monteiro, J.M., Rao, A., Shawe-Taylor, J., Mourão-Miranda, J., 2016. A multiple hold-out framework for sparse partial least squares. Journal of Neuroscience Methods 271, 182–194.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825–2830.

Petersen, R., Aisen, P., Beckett, L., Donohue, M., Gamst, A., Harvey, D., Jack, C., Jagust, W., Shaw, L., Toga, A., et al., 2010. Alzheimer's disease neuroimaging initiative (adni) clinical characterization. Neurology 74, 201–209.

Rahim, M., Thirion, B., Comtat, C., Varoquaux, G., 2016. Transmodal learning of functional networks for alzheimer's disease prediction. IEEE Journal of Selected Topics in Signal Processing 10, 1204–1213.

Rashid, B., Arbabshirani, M.R., Damaraju, E., Cetin, M.S., Miller, R., Pearlson, G.D., Calhoun, V.D., 2016. Classification of schizophrenia and bipolar patients using static and dynamic resting-state fMRI brain connectivity. NeuroImage 134, 645–657.

Savio, A., Graña, M., 2015. Local activity features for computer aided diagno-

sis of schizophrenia on resting-state fMRI. Neurocomputing 164, 154–161.
Schrouff, J., Mourão-Miranda, J., Phillips, C., Parvizi, J., 2016. Decoding intracranial EEG data with multiple kernel learning method. Journal of Neuroscience Methods 261, 19–28.
Segal, M., Xiao, Y., 2011. Multivariate random forests. WIREs Data Mining Knowl Discov 1, 80–87.
Shaw, L.M., Vanderstichele, H., Knapik-Czajka, M., Figurski, M., et al., 2011. Qualification of the analytical and clinical performance of CSF biomarker analyses in ADNI. Acta Neuropathologica 121.
Smith, S.M., Nichols, T.E., Vidaurre, D., Winkler, A.M., Behrens, T.E.J., Glasser, M.F., Ugurbil, K., Barch, D.M., Essen, D.C.V., Miller, K.L., 2015. A positive-negative mode of population covariation links brain connectivity, demographics and behavior. Nature Neuroscience 18, 1565–1567.
Sui, J., Adali, T., Yu, Q., Chen, J., Calhoun, V.D., 2012. A review of multivariate methods for multimodal fusion of brain imaging data. Journal of Neuroscience Methods 204, 68–81.
Thompson, P.M., Stein, J.L., Medland, S.E., Hibar, D.P., Vasquez, A.A., Renteria, M.E., Toro, R., Jahanshad, N., Schumann, G., Franke, B., et al., 2014. The enigma consortium: large-scale collaborative analyses of neuroimaging and genetic data. Brain imaging and behavior 8, 153–182.
Van Essen, D.C., Smith, et al., 2013. The wu-minn human connectome project: an overview. Neuroimage 80, 62–79.
Varoquaux, G., Raamana, P.R., Engemann, D.A., Hoyos-Idrobo, A., Schwartz, Y., Thirion, B., 2016. Assessing and tuning brain decoders: cross-validation, caveats, and guidelines. NeuroImage .
Vounou, M., Nichols, T.E., Montana, G., Initiative, A.D.N., et al., 2010. Discovering genetic associations with high-dimensional neuroimaging phenotypes: a sparse reduced-rank regression approach. Neuroimage 53.
Wang, H., Nie, F., Huang, H., Risacher, S., Ding, C., Saykin, A.J., Shen, L., ADNI, 2011. Sparse multi-task regression and feature selection to identify brain imaging predictors for memory performance, in: 2011 International Conference on Computer Vision, Institute of Electrical and Electronics Engineers (IEEE).
Wang, H., Nie, F., Huang, H., Yan, J., Kim, S., Risacher, S., Saykin, A., Shen, L., 2012. High-order multi-task feature learning to identify longitudinal phenotypic markers for alzheimer's disease progression prediction, in: NIPS, pp. 1277–1285.
Wang, L., Zang, Y., He, Y., Liang, M., Zhang, X., Tian, L., Wu, T., Jiang, T., Li, K., 2006. Changes in hippocampal connectivity in the early stages of alzheimerś disease: Evidence from resting state fMRI. NeuroImage 31, 496–504.
Wang, X., Zhang, T., Chaim, T.M., Zanetti, M.V., Davatzikos, C., 2015. Classification of MRI under the presence of disease heterogeneity using multi-task learning: Application to bipolar disorder, in: MICCAI 2015, pp. 125–132.
Woo, C.W., Chang, L.J., Lindquist, M.A., Wager, T.D., 2017. Building better biomarkers: brain models in translational neuroimaging. Nature Neuroscience 20, 365–377.
Yuan, L., Wang, Y., Thompson, P.M., Narayan, V.A., Ye, J., 2012. Multi-source feature learning for joint analysis of incomplete multiple heterogeneous neuroimaging data. NeuroImage 61, 622–632.
Yuan, M., Lin, Y., 2006. Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68, 49.
Zhang, D., Shen, D., et al., 2012. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in alzheimer's disease. NeuroImage 59, 895–907.